The AL Basis for the solution of elliptic problems in heterogeneous media

Grasedyk, L ; Greff, I ; Sauter, S

Abstract: In this paper, we will show that, for elliptic problems in heterogeneous media, there exists a local (generalized) finite element basis (AL basis) consisting of O((log 1/H)(d+1) basis functions per nodal point such that the convergence rates of the classical finite element method for Poisson- type problems are preserved. Here H denotes the mesh width of the finite element mesh and d is the spatial dimension. We provide several numerical examples beyond our theory, where even O(1) basis functions per nodal point are sufficient to preserve the convergence rates.

DOI: https://doi.org/10.1137/11082138X

MULTISCALE MODEL. SIMUL.
Vol. 10, No. 1, pp. 245–258

THE AL BASIS FOR THE SOLUTION OF ELLIPTIC PROBLEMS IN HETEROGENEOUS MEDIA*

L. GRASEDYCK†, I. GREFF‡, AND S. SAUTER§

Abstract. In this paper, we will show that, for elliptic problems in heterogeneous media, there exists a local (generalized) finite element basis (AL basis) consisting of $O\big(\big(\log \frac{1}{H}\big)^{d+1}\big)$ basis functions per nodal point such that the convergence rates of the classical finite element method for Poisson-type problems are preserved. Here H denotes the mesh width of the finite element mesh and d is the spatial dimension. We provide several numerical examples beyond our theory, where even $O(1)$ basis functions per nodal point are sufficient to preserve the convergence rates.

Key words. elliptic problem, heterogeneous media, Green's function, generalized finite elements

AMS subject classifications. 65N30, 35B65, 35J57

DOI. 10.1137/11082138X

1. Introduction.

The efficient numerical modeling of elliptic problems in heterogeneous media is of fundamental and practical importance and arises in applications such as composite materials, porous media, and turbulent transport. If the geometric details, e.g., inclusions in the material, have complicated structure and/or are tiny, then the resolution of all details by conventional finite elements becomes too costly—especially for three-dimensional problems.

In recent years, many types of *generalized* finite element methods have been developed where the characteristic physical behavior of the solution is incorporated in the *shape* of the trial functions so that the geometric details may not be resolved by the finite element mesh, while the goal is to preserve the asymptotic convergence rates also for these coarse-scale discretizations. Early papers on this topic are [2], [3]. We omit an extensive list of references on the construction of generalized finite element methods because our goal here is more theoretical and concerned with the following problem: Consider an elliptic Poisson-type problem with sufficiently smooth diffusion coefficients, right-hand side, and domain boundary. Then a discretization with continuous piecewise linear finite elements on a finite element mesh with mesh width H converges with respect to the energy norm at a rate of $O(H)$. If the diffusion coefficient is nonsmooth as is typical for heterogeneous media it is well known that the convergence rate becomes very poor. In this paper, we will address the question of whether the linear finite element space can be enriched by "a few" additional shape functions so that the convergence rate is $O(H)$ without any regularity assumption on the diffusion coefficient.

A similar question was also addressed in the recent paper [4]. There, the local finite element spaces are constructed as the span of the solutions of local eigenvalue problems, while in our approach the local finite element spaces are the L^2-projected

*Received by the editors January 17, 2011; accepted for publication (in revised form) January 9, 2012; published electronically March 13, 2012.
http://www.siam.org/journals/mms/10-1/82138.html

†Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, D-52056 Aachen, Germany (lgr@igpm.rwth-aachen.de).

‡Laboratoire de Mathématiques Appliquées, Université de Pau et des pays de l'Adour, F-64013 Pau Cedex, France (isabelle.greff@univ-pau.fr).

§Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland (stas@math.uzh.ch).

localized images of the standard finite element basis under the inverse differential operator. The advantage of the method in [4] is that the generalized finite element basis functions are constructed in one stroke, while in our approach a local, higher dimensional, auxiliary space is first generated and then projected to a lower dimensional space. On the other hand, our approach requires only the solution of the PDE for a localized right-hand side (which can be efficiently performed, e.g., with an algebraic multigrid algorithm), while the method in [4] requires the solution of eigenvalue problems. Furthermore, our method results in a generalized finite element space which can be applied to any right-hand side $f \in L^2(\Omega)$ without any modification. The approach in [4] requires the construction of a particular local solution by solving certain Neumann problems for every given right-hand side in a preprocessing step. Further approaches for the construction and analysis of a multiscale basis for problems with high contrast (without periodicity assumptions) include [15], [13], [16].

In order to make a fair comparison of both methods, a fully discrete and fast version of both algorithms has to be developed, which will be a topic of future research.

Another related approach (see [1]) is based on the proper orthogonal decomposition (POD) technique with the goal of generating low-dimensional subspaces that contain the essential information of the solution. There the conventional finite element space is enriched by the solutions of certain eigenvalue problems—a snapshot technique is used in order to reduce the dimension of these eigenvalue problems. Numerical tests for Helmholtz-type problems and problems in heat transfer show good convergence behavior, while an error analysis which is explicit in the problem coefficients is still open. In contrast, our approach *guarantees* that the AL basis converges with a rate of $O(H)$ for every right-hand side in L^2.

We have omitted a discussion of an *hp-version* of our method in order not to overload the paper, but we focus on the question of whether a linear finite element space—enriched by a few generalized basis functions—can recover the linear convergence rate.

2. Model problem. In this paper we shall deal with the following problem: Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain, and let the diffusion matrix $A \in L^\infty\left(\Omega, \mathbb{R}^{d\times d}_{\mathrm{sym}}\right)$ be uniformly elliptic:

$$0 < \alpha = \operatorname*{essinf}_{x\in\Omega} \inf_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle Av, v\rangle}{\langle v, v\rangle} \;\le\; \operatorname*{esssup}_{x\in\Omega} \sup_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle Av, v\rangle}{\langle v, v\rangle} = \beta < \infty. \tag{1}$$

For $m \in \mathbb{N}_0$, let $H^m(\Omega)$ denote the usual Sobolev space with norm $\|\cdot\|_{H^m(\Omega)}$, and let $H^m_0(\Omega)$ be the closure of $C^\infty_0(\Omega)$ with respect to the norm $\|\cdot\|_{H^m(\Omega)}$. The dual space of $H^m_0(\Omega)$ is denoted by $H^{-m}(\Omega)$.

For given $f \in L^2(\Omega)$, we are seeking $u \in H^1_0(\Omega)$ such that

$$a(u,v) := \int_\Omega \langle A\nabla u, \nabla v\rangle = \int_\Omega fv =: F(v) \qquad \forall v \in H^1_0(\Omega). \tag{2}$$

The abstract conforming Galerkin method to this problem is given by specifying a finite-dimensional subspace $S \subset H^1_0(\Omega)$ and seeking $u_S \in S$ such that

$$a(u_S, v) = F(v) \qquad \forall v \in S. \tag{3}$$

In order to avoid technicalities with curved finite elements, we assume that Ω is either a one-dimensional interval, or a two-dimensional polygonal domain, or a three-dimensional polyhedron. Let $\mathcal{G} = \{\tau_i : 1 \le i \le N\}$ be a regular finite element mesh

in the sense of Ciarlet [7]. More precisely, $\mathcal{G}$ is a partition of Ω into d-dimensional disjoint open simplices which satisfy the following:

a. For any two nonidentical elements $\tau, t \in \mathcal{G}$, the intersection $\overline{\tau} \cap \overline{t}$ either is empty, a common vertex (interval endpoint in one dimension), a common edge (for $d \geq 2$), or a common face (for $d = 3$).
b. $\overline{\Omega} = \bigcup_{\tau \in \mathcal{G}} \overline{\tau}$.

The mesh width is given by $H = \max \{\operatorname{diam} \tau : \tau \in \mathcal{G}\}$. The shape regularity of $\mathcal{G}$ is described by the constant

$$\kappa := \max \left\{ \frac{\operatorname{diam} \tau}{\rho_\tau} : \tau \in \mathcal{G} \right\},$$

where ρ_τ is the diameter of the maximal inscribed ball in τ. Since $\mathcal{G}$ contains finitely many simplices, the constant κ is always bounded but becomes large if the simplices are degenerate, e.g., are flat or needle-shaped. The constants in the following estimates depend on the mesh via the constant κ—they are bounded for any fixed κ but, possibly, become large for large κ.

The space of continuous, piecewise linear finite elements is given by

$$S = \left\{ u \in H_0^1(\Omega) \mid \forall \tau \in \mathcal{G} : u_\tau \in \mathbb{P}_1 \right\}. \tag{4}$$

Let $(b_i)_{i=1}^n$ denote the usual local nodal basis of S ("hat functions"); their support is denoted by

$$\omega_i := \operatorname{supp} b_i. \tag{5}$$

The shape regularity of $\mathcal{G}$ implies the local quasi-uniformity of the mesh as follows. We define simplex layers around ω_i by the following recursion: Let $\omega_{i,0} := \omega_i$, and define, for $j = 0, 1, 2, \ldots$,

$$\omega_{i,j+1} := \bigcup \{ \bar{\tau} \mid \tau \in \mathcal{G} \text{ and } \omega_{i,j} \cap \bar{\tau} \neq \emptyset \}. \tag{6}$$

Then, for all $1 \leq i \leq n$ and $m \in \mathbb{N}_0$, there exists a constant $c_{m,\kappa}$ depending only on m and κ such that

$$\rho_\tau \geq c_{m,\kappa} \operatorname{diam} t \qquad \forall \tau, t \in \omega_{i,m}. \tag{7}$$

If the data of the continuous problem (2), i.e., the diffusion coefficient A, the right-hand side f, and the domain Ω, are sufficiently smooth so that the solution is in $H^2(\Omega) \cap H_0^1(\Omega)$, then the Galerkin discretization (3) based on the finite element space S (as in (4) with a shape regular simplicial finite element mesh $\mathcal{G}$) has a unique solution u_S which satisfies the error estimate

$$\|u - u_S\|_{H^1(\Omega)} \leq CH \|f\|_{L^2(\Omega)}.$$

The details can be found in any textbook on the numerical analysis of the finite element method, and we refer the reader to, e.g., [7]. We denote this linear convergence with respect to H as the "textbook" convergence rate of linear finite elements. It is well known that—as long as the mesh $\mathcal{G}$ does not resolve the (possible) discontinuities and oscillations of A—the textbook-convergence rates of linear finite elements are substantially reduced.

In this paper we will address the following question: Is there a set of basis functions $b_{i,j} \in H_0^1(\Omega)$, $1 \le j \le p$, $1 \le i \le n := \dim(S)$, such that

$$\operatorname{supp} b_{i,j} \subset \omega_i$$

and the *linear convergence property* (cf. Definition 1) holds?

DEFINITION 1 (linear convergence property). *Let $a(\cdot,\cdot)$ be as in* (2)*, and let S be as in* (4)*, with supports ω_i of basis functions as in* (5)*. Let $\tilde{S} \subset H_0^1(\Omega)$ be a finite-dimensional subspace which satisfies*

$$\tilde{S} = \operatorname{span}\{b_{i,j} \mid 1 \le j \le p, 1 \le i \le n \text{ and } \operatorname{supp} b_{i,j} \subset \omega_i\}. \tag{8}$$

$\tilde{S}$ has the linear convergence property (LCP) *if, for any $f \in L^2(\Omega)$, the solution to the problem of finding $u_{\tilde{S}} \in \tilde{S}$ such that*

$$a(u_{\tilde{S}}, v) = \int_\Omega fv \qquad \forall v \in \tilde{S}$$

satisfies the error estimate

$$\|u - u_{\tilde{S}}\|_{H^1(\Omega)} \le CH \|f\|_{L^2(\Omega)},$$

where C depends only on α and β (cf. (1)*).*

Remark 2. Note that the linear convergence property is defined for a *given* set of supports ω_i. More generally, one could also include an optimal choice of these supports in the definition. This would be appropriate if the (possibly low) regularity of the solution is *not* distributed uniformly over the domain. Our simplified definition is suitable for problems where the diffusion coefficient is rough/oscillating over the whole domain and a quasi-uniform mesh $\mathcal{G}$ is an adequate choice.

We will prove theoretically that, by choosing the number p in (8) proportionally to $O\big(\log^{d+1}\frac{1}{H}\big)$, such a set of basis functions exists. On the other hand, we will check the optimality of this result by constructing an alternative basis based on the singular value decomposition of an overrefined discretization of the problem—this approach, conceptually, is not suited for practical computations because of its prohibitively high numerical cost. However, the experiments show that, for some diffusion matrices with very complicated oscillations, it might be possible that only $O(1)$ basis functions per nodal point are sufficient to preserve the linear convergence property.

We emphasize that our construction of $b_{i,j}$ is only semidiscrete, and we consider our results at this stage as a theoretical insight rather than a practical method. Forthcoming papers will address the question of how to construct the basis $b_{i,j}$ efficiently.

For problems with periodic coefficient $A = A_0\left(\frac{\cdot}{\varepsilon}\right)$ or locally periodic coefficient $A(\cdot) = A_0\left(\cdot, \frac{\cdot}{\varepsilon}\right)$ (with slowly varying functions $A_0(\cdot)$, resp., $A_0(\cdot,\cdot)$) the convergence of finite elements and generalized finite elements, as well as methods based on homogenization-type techniques, has been analyzed in the literature (see, e.g., [14], [10], and for a more general, nonperiodic setting see [17]). In contrast, we do not impose any assumption on the structure of A but only assume uniform ellipticity and continuity for the corresponding bilinear form. Our results rely strongly on the approximability of the Green's function for general elliptic problems (see [5]).

3. The AL basis. Before we describe the construction in detail we first sketch the idea. The definition of the AL basis is patchwise. For a patch ω_i (cf. (5)), the set of indices $\mathcal{I} := \{1, 2, \ldots, n\}$ is split into a nearfield $\mathcal{I}_i^{\text{near}}$, which contains those

indices j which correspond to basis functions with support close to ω_i, and into the remaining index set $\mathcal{I}_i^{\text{far}}$ which represents the farfield. For the index $i \in \mathcal{I}$, one part of the AL basis is given by $b_i L^{-1} b_j$, $j \in \mathcal{I}_i^{\text{near}}$, where $L : H_0^1(\Omega) \to H^{-1}(\Omega)$ denotes the operator associated with the bilinear form $a(\cdot,\cdot)$, and b_i is the standard finite element basis (cf. (5)).

For the other part, we first set up an auxiliary space $X_i^{\text{far}} := \operatorname{span}\left\{ L^{-1} b_j \big|_{\omega_i^\star} : j \in \mathcal{I}_i^{\text{far}} \right\}$ in a certain neighborhood $\omega_i^\star \supset \omega_i$. The space X_i^{far} allows for a low-dimensional approximation which is constructed in the second step: Introduce intermediate neighborhoods $\omega_i = D_{i,\ell} \subset D_{i,\ell-1} \subset \cdots \subset D_{i,0} \subset \omega_i^\star$, where $\ell = O\left(\log \frac{1}{H}\right)$. For any $D_{i,j}$, a mesh $\mathcal{G}_{i,j}$ is constructed by intersecting $D_{i,j}$ with a regular Cartesian mesh of width $O\left(H/\log \frac{1}{H}\right)$.

Then the farfield part of the AL basis for the patch ω_i is given by $b_i P \chi_\tau$, $\tau \in \mathcal{G}_{i,j}$, $0 \le j \le \ell$, where P is the L^2-orthogonal projection on X_i^{far}, and χ_τ denotes the characteristic function for $\tau \in \mathcal{G}_{i,j}$.

Now we come to the detailed description of the construction. Let

$$B_i := L^{-1} b_i, \qquad i \in \mathcal{I} := \{1, 2, \ldots, n\}. \tag{9}$$

Remark 3. The condition $\operatorname{supp} b_{i,j} \subset \omega_i$ in (8) on the localness of the basis functions implies the sparsity of the arising stiffness matrix and is crucial for the computational/storage complexity of the discretization. Without this condition, the basis B_i would be a very good choice for preserving the optimal error estimates. However, the functions B_i, in general, are nonlocal, and the generation of the system matrix would be prohibitively expensive.

Recall the definition of the simplex layers $\omega_{i,j}$ as in (6). Depending on the parameter $\eta \in (0,1)$ we define $\omega_i^\star := \omega_{i,m}$, where m is chosen such that

$$\eta \operatorname{diam} \omega_i \le \operatorname{dist}(\omega_i, \partial\omega_i^\star). \tag{10}$$

Due to the local quasi-uniformity of the mesh (cf. (7)) we can choose $0 < \eta$ sufficiently small but independent of H such that $m = m(\eta) = O(1)$. Note that the constant C in the approximation property (see (19)) depends on η but is independent of H. To reduce technicalities we assume that, for all $1 \le i \le m$, the sets ω_i and $\omega_i^\star$ are convex.

For an index $i \in \mathcal{I}$, we define a nearfield and a farfield by setting[1]

$$\mathcal{I}_i^{\text{near}} := \{ j \in \mathcal{I} \mid \ 0 < |\omega_i^\star \cap \operatorname{supp} b_j| \} \quad \text{and} \quad \mathcal{I}_i^{\text{far}} := \mathcal{I} \backslash \mathcal{I}_i^{\text{near}}.$$

Then, we set

$$X_i^{\text{far}} := \operatorname{span}\left\{ B_j \big|_{\omega_i^\star} \mid \ j \in \mathcal{I}_i^{\text{far}} \right\} \tag{11}$$

and

$$V_i^{\text{near}} := \operatorname{span}\{ b_i B_j \mid \ j \in \mathcal{I}_i^{\text{near}} \}. \tag{12}$$

Note that the functions in X_i^{far} are L-harmonic in $\omega_i^\star$; i.e., any $v \in X_i^{\text{far}}$ satisfies

$$\int_{\omega_i^\star} \langle A \nabla v, \nabla w \rangle = 0 \qquad \forall w \in H_0^1(\omega_i^\star).$$

[1] For a measurable subset $M \subset \mathbb{R}^d$, we set $|M| := \int_M 1$.

It turns out that the space X_i^{far} can be approximated by a low-dimensional space, and we employ the construction proposed in [5].

We introduce intermediate layers between ω_i and $\omega_i^\star$ by setting $r_{i,0} := \operatorname{dist}(\omega_i, \partial\omega_i^\star)$ and

$$r_{i,j} := \left(1 - \frac{j}{\ell}\right) r_{i,0}, \qquad 1 \le j \le \ell,$$

where ℓ will be fixed later. The intermediate layers are given by

$$D_{i,j} := \{x \in \omega_i^\star \mid \operatorname{dist}(x, \omega_i) \le r_{i,j}\}, \qquad 0 \le j \le \ell,$$

and satisfy $\omega_i = D_{i,\ell} \subset D_{i,\ell-1} \subset \cdots \subset D_{i,0} \subset \omega_i^\star$. For $\rho > 0$, let $\mathcal{G}_\rho$ denote a Cartesian tensor mesh on $\mathbb{R}^d$ consisting of d-dimensional elements (hypercubes) with side lengths ρ, and let

$$\mathcal{G}_{i,j} := \left\{ D_{i,j} \cap \tau \;\middle|\; \tau \in \mathcal{G}_\rho,\ \rho := \frac{\operatorname{diam} D_{i,j}}{k} \right\},$$

where $k \in \mathbb{N}_{\ge 1}$ will be fixed later. For $t \in \mathcal{G}_{i,j}$, we denote the characteristic function for t by $\chi_t : \Omega \to \mathbb{R}$. We define

$$\widetilde{V}_{i,j}^{\mathrm{far}} := \operatorname{span}\left\{ (P\chi_t)|_{\omega_i} \mid t \in \mathcal{G}_{i,j} \right\},$$

where $P : L^2(\Omega) \to X_i^{\mathrm{far}}$ is the L^2-orthogonal projection. Then,

$$\widetilde{V}_i^{\mathrm{far}} := \widetilde{V}_{i,0}^{\mathrm{far}} + \widetilde{V}_{i,1}^{\mathrm{far}} + \cdots + \widetilde{V}_{i,\ell}^{\mathrm{far}} \tag{13}$$

and, finally,

$$V_i^{\mathrm{far}} := \left\{ b_i v \mid v \in \widetilde{V}_i^{\mathrm{far}} \right\}. \tag{14}$$

Since $b_i \in W_0^{1,\infty}(\Omega)$ and $X_i^{\mathrm{far}} \subset H^1(\omega_i^\star)$, we conclude that $b_i v \in H_0^1(\omega_i)$ for all $v \in \widetilde{V}_i^{\mathrm{far}}$. Thus, we can identify $b_i v$ by its extension by zero to a function (again denoted by $b_i v$) in $H_0^1(\Omega)$. In this sense, we have

$$V_i^{\mathrm{far}} \subset H_0^1(\Omega), \qquad \dim V_i^{\mathrm{far}} \le \sum_{j=0}^{\ell} \#\mathcal{G}_{i,j} \le \sum_{j=0}^{\ell} k^d = (\ell+1)\, k^d.$$

Definition 4 (AL basis). *For any support ω_i (cf. (5)) the set of AL basis functions consists of the functions $b_i B_j$, $j \in \mathcal{I}_i^{\mathrm{near}}$, and the functions*

$$b_i P\chi_t \qquad \forall t \in \mathcal{G}_{i,q}, \qquad 0 \le q \le \ell.$$

The general notation is $b_{i,j}$, $1 \le j \le p$, $1 \le i \le n$, where $p := \dim\left(V_i^{\mathrm{far}} + V_i^{\mathrm{near}}\right)$. The corresponding generalized finite element space is given by

$$V_{\mathrm{AL}} := \left(V_1^{\mathrm{near}} + V_1^{\mathrm{far}}\right) + \left(V_2^{\mathrm{near}} + V_2^{\mathrm{far}}\right) + \cdots + \left(V_n^{\mathrm{near}} + V_n^{\mathrm{far}}\right). \tag{15}$$

Remark 5. Since the index m in the definition of $\omega_i^\star$ is independent of H, we have $\dim V_i^{\mathrm{near}} = O(1)$. As a consequence of the error analysis, it will turn out that $\dim V_i^{\mathrm{far}} = O\left(\log^{d+1} \frac{1}{H}\right)$.

The Galerkin discretization for the generalized finite element space V_{AL} is given by seeking $u_{\mathrm{AL}} \in V_{\mathrm{AL}}$ such that

$$a(u_{\mathrm{AL}}, v) = F(v) \qquad \forall v \in V_{\mathrm{AL}}. \tag{16}$$

4. Error analysis. The error analysis is based on the results in [5]. We know that the constants in the error estimates of this section depend on α and $\beta \in \mathbb{R}_{>0}$ without writing this dependence explicitly. Our emphasis is on proving that the estimates are uniform for all diffusion matrices $A \in L^\infty\left(\Omega, \mathbb{R}^{d\times d}_{\mathrm{sym}}\right)$ which satisfy (1). Note that the assumptions on A imply

$$\left\|L^{-1}\right\|_{H^1_0(\Omega)\leftarrow H^{-1}(\Omega)} \leq C.$$

ASSUMPTION 6. *The domains ω_i, $\omega_i^\star$ (cf. (5) and (10)) are convex and satisfy (10) for some $\eta \gtrsim 1$. The constant*

$$C_\sharp := \max_{i\in\mathcal{I}} \#\mathcal{I}_i^{\mathrm{near}}$$

depends only on the shape-regularity of the finite element mesh $\mathcal{G}$ and the number $m = O(1)$ (depending on the local *quasi-uniformity of $\mathcal{G}$) in the definition of ω_i. Finally, there exists a constant C_q such that*

$$\#\mathcal{I} \leq C_q H^{-d}.$$

THEOREM 7. *Let u denote the solution of (2). Let the parameters ℓ and k in the definition of the farfield part of V_{AL} be chosen according to*

$$\ell := \max\left\{2, \left\lceil \frac{2+d}{2\log 2}\log\frac{1}{H}\right\rceil\right\} \quad \text{and} \quad k := \left\lceil \frac{2c_0\ell^2}{(\ell-1)}\right\rceil \tag{17}$$

for some $c_0 = O(1)$. Let u_{AL} be the corresponding Galerkin solution (cf. (16)). Then, the error estimate

$$\|u - u_{\mathrm{AL}}\|_{H^1(\Omega)} \leq CH\|f\|_{L^2(\Omega)}$$

holds while

$$\dim V_{\mathrm{AL}} \leq C_d N \ell^{d+1} \leq \tilde{C}_d H^{-d}\log^{d+1}\frac{1}{H}.$$

Proof. Let $P_S : L^2(\Omega) \to S$ denote the L^2-orthogonal projection onto S. For $f \in L^2(\Omega)$, let $u = L^{-1}f$. Then, the substitution of f by $P_S f$ leads to a consistent perturbation

$$\left\|u - L^{-1}P_S f\right\|_{H^1(\Omega)} \leq C\|f - P_S f\|_{H^{-1}(\Omega)} \leq CH\|f\|_{L^2(\Omega)}. \tag{18}$$

We introduce the nearfield and the farfield parts of f with respect to some $i \in \mathcal{I}$ by

$$f_i^{\mathrm{near}} := \sum_{j\in\mathcal{I}_i^{\mathrm{near}}} (P_S f)_j\, b_j \quad \text{and} \quad f_i^{\mathrm{far}} := \sum_{j\in\mathcal{I}_i^{\mathrm{far}}} (P_S f)_j\, b_j,$$

where $(P_S f)_j := (P_S f)(x_j)$ and x_j is the nodal point corresponding to b_j. Then,

$$L^{-1}P_S f = \sum_{i=1}^n \underbrace{b_i L^{-1} f_i^{\mathrm{near}}}_{u_i^{\mathrm{near}}} + \sum_{i=1}^n b_i \underbrace{L^{-1} f_i^{\mathrm{far}}}_{u_i^{\mathrm{far}}}.$$

Since $u_i^{\mathrm{near}} \in V_i^{\mathrm{near}}$, the approximation problem is reduced to the approximation of u_i^{far}. Note that the function $u_i^{\mathrm{far}}\big|_{\omega_i^\star} \in X_i^{\mathrm{far}}$. The following approximation estimate is based on the results in [5]: There exists $\tilde{u}_i^{\mathrm{far}} \in \widetilde{V}_i^{\mathrm{far}}$ (cf. (13)) such that

$$\left\|u_i^{\mathrm{far}} - \tilde{u}_i^{\mathrm{far}}\right\|_{H^m(\omega_i)} \leq CH^{s-m} \left\|\nabla L^{-1} f_i^{\mathrm{far}}\right\|_{L^2(\omega_i)}, \qquad m = 0, 1, \tag{19}$$

with $s = 2 + d/2$. To see this we argue as in [6, second to last estimate p. 172] by choosing $p \leftarrow \ell$ therein (ℓ is defined in (17)) to obtain the estimate in the H^1-seminorm. For the L^2-norm we use [6, second estimate, p. 172] with $i \leftarrow \ell$ and $\delta \leftarrow O(H)$. The approximation of u finally is given by

$$\tilde{u} := \sum_{i=1}^n u_i^{\mathrm{near}} + \sum_{i=1}^n b_i \tilde{u}_i^{\mathrm{far}} \in V_{\mathrm{AL}}.$$

By using (18) and a triangle inequality we obtain

$$\|u - \tilde{u}\|_{H^1(\Omega)} \leq CH \|f\|_{L^2(\Omega)} + \left\|\sum_{i=1}^n b_i \left(u_i^{\mathrm{far}} - \tilde{u}_i^{\mathrm{far}}\right)\right\|_{H^1(\Omega)}.$$

The second sum can be estimated by combining the Leibniz rule for products with a triangle inequality, a Hölder's inequality, an inverse inequality for b_i, and (19):

$$\begin{aligned}
\left\|\sum_{i=1}^n b_i \left(u_i^{\mathrm{far}} - \tilde{u}_i^{\mathrm{far}}\right)\right\|_{H^1(\Omega)} &\leq \sum_{i=1}^n \Big(\|b_i\|_{L^\infty(\omega_i)} \left\|u_i^{\mathrm{far}} - \tilde{u}_i^{\mathrm{far}}\right\|_{H^1(\omega_i)} \\
&\qquad + \|\nabla b_i\|_{L^\infty(\omega_i)} \left\|u_i^{\mathrm{far}} - \tilde{u}_i^{\mathrm{far}}\right\|_{L^2(\omega_i)} \Big) \\
&\leq CH^{s-1} \sum_{i=1}^n \left\|\nabla L^{-1} f_i^{\mathrm{far}}\right\|_{L^2(\omega_i)} \\
&\leq CH^{s-1} \sum_{i=1}^n \left(\left\|\nabla L^{-1} P_S f\right\|_{L^2(\omega_i)} + \left\|\nabla L^{-1} f_i^{\mathrm{near}}\right\|_{L^2(\omega_i)} \right) \\
&\leq CH^{s-1} \sqrt{n} \left(\|f\|_{L^2(\Omega)} + \sqrt{\sum_{i=1}^n \|\nabla L^{-1} f_i^{\mathrm{near}}\|_{L^2(\omega_i)}^2} \right).
\end{aligned}$$

In order to estimate the last sum we use the representation of L^{-1} via the Green's function

$$L^{-1} f_i^{\mathrm{near}} = \int_\Omega G(x, y) f_i^{\mathrm{near}}(y)\, dy,$$

where the estimate

$$\sup_{x \in \Omega} \|\nabla G(x, y)\|_{L^\rho(\Omega)} \leq C_{d,\alpha,\beta,\varepsilon} \quad \text{with} \quad \alpha, \beta \text{ as in (1)}, \quad \rho := \frac{d}{d-1} - \varepsilon,$$

for any $0 < \varepsilon \leq \frac{1}{d-1}$ follows from [11, Theorem 1.1 and (1.12)] for $d \geq 3$ and from [9, Remark 2.19] for $d = 2$. For $d = 1$ the estimate

$$\sup_{x \in \Omega} \|G'(x, y)\|_{L^\infty(\Omega)} \leq C_{\alpha,\beta}$$

follows from [12, (10.14)]. In the following we work out only the case $d \geq 2$, while the case $d = 1$ can be derived analogously. Hence,

$$\begin{aligned}\left\|\nabla L^{-1} f_i^{\text{near}}\right\|_{L^2(\omega_i)}^2 &\leq \int_{\omega_i} \left| \int_\Omega \nabla_x G(x,y)\, f_i^{\text{near}}(y)\, dy \right|^2 dx \\ &\leq C_{d,\alpha,\beta,\varepsilon}^2 \left|\omega_i\right| \left\| f_i^{\text{near}} \right\|_{L^p(\Omega)}^2 \leq C_{d,\alpha,\beta,\varepsilon}^2 H^d \left\| f_i^{\text{near}} \right\|_{L^p(\Omega)}^2\end{aligned}$$

for $p = \frac{d+\varepsilon(1-d)}{1+\varepsilon(1-d)} \geq 2$. From [8, Proposition 3.10] (choosing $p' \leftarrow p$, $p \leftarrow 2$, $\alpha \leftarrow 0$ therein) we conclude that

$$\left\| f_i^{\text{near}} \right\|_{L^p(\Omega)}^2 \leq H^{-\zeta} \left\| f_i^{\text{near}} \right\|_{L^2(\Omega)}^2, \qquad \zeta := \frac{d\,(d - \varepsilon\,(1-d) - 2)}{d + \varepsilon\,(1-d)},$$

so that

$$\left\|\nabla L^{-1} f_i^{\text{near}}\right\|_{L^2(\omega_i)}^2 \leq C_{d,\alpha,\beta,\varepsilon}^2 H^{2-2q} \left\| f_i^{\text{near}} \right\|_{L^2\left(\omega_i^\star\right)}^2 \leq C_{d,\alpha,\beta,\varepsilon}^2 H^{2-2q} \left\| P_S f \right\|_{L^2\left(\omega_i^\star\right)}^2,$$

where $q := \varepsilon \frac{(d-1)^2}{d-\varepsilon(d-1)}$. Hence,

$$\sum_{i=1}^n \left\|\nabla L^{-1} f_i^{\text{near}}\right\|_{L^2(\omega_i)}^2 \leq C_{d,\alpha,\beta,\varepsilon}^2 H^{2-2q} \sum_{i=1}^n \left\| P_S f \right\|_{L^2\left(\omega_i^\star\right)}^2 \leq C_{d,\alpha,\beta,\varepsilon}^2 H^{2-2q} \left\| f \right\|_{L^2(\Omega)}^2.$$

In summary, we have proved (by using $q \leq 1$ for all $\varepsilon \in [0, 1/(d-1)]$)

$$\begin{aligned}\left\| u - \tilde{u} \right\|_{H^1(\Omega)} &\leq CH \left\| f \right\|_{L^2(\Omega)} + CH^{s-1} \sqrt{n} \left(\left\| f \right\|_{L^2(\Omega)} + H^{1-q} \left\| f \right\|_{L^2(\Omega)} \right) \\ &\overset{\text{Assumpt. 6}}{\leq} C \left(H + \sqrt{C_q} H^{s-1-d/2} \right) \left\| f \right\|_{L^2(\Omega)},\end{aligned}$$

and the choice of s yields the assertion. ∎

5. On the optimality of the AL basis. Our theory shows that $O\left(\log^{d+1} \frac{1}{H}\right)$ AL basis functions per nodal point are sufficient to preserve the linear approximation property. In this section we will investigate whether this number of basis functions is optimal or whether, for an "optimally chosen" basis, only $O(1)$ basis functions per nodal point is sufficient for some problems with a complicated diffusion matrix in order to preserve the linear convergence property. We emphasize that the AL basis at this stage is only semidiscrete—a forthcoming paper will be concerned with a numerical realization of some approximate version of the AL basis.

5.1. Locally optimal subspaces. The choice of experiments is based on some heuristics: As a subset of all possible right-hand sides we consider real valued plane waves of the form

$$f_j := \sin(2\pi \langle \xi_j, \cdot \rangle), \quad \xi_j = \left(\sin \frac{\pi j}{20}, \cos \frac{\pi j}{20} \right)^\intercal, \quad j = 1, \ldots, 20, \tag{20}$$

and the computational domain $\Omega := (-1,1) \times (-1,1)$.

We expect (and observe) that these choices of right-hand sides generate solutions which exhibit very different directions of oscillations so that the use of only one (generalized) shape function per nodal point is a critical test for approximability. As test examples we have considered the following diffusion matrices.

Problem 1. In this case, the diffusion coefficient is oscillatory:

$$A_1(x,y) = \nu(x,y)I \quad \text{with} \quad \nu(x,y) := 2 - \frac{\cos\left(\frac{2\pi x^2}{\varepsilon}\right) + \cos\left(\frac{2\pi y}{\varepsilon}\right)}{11\left(\frac{1}{10} + \left(x+y\right)^2\right)} \tag{21}$$

and I denotes the 2×2 identity matrix. Note that this coefficient satisfies assumption (1) with $\alpha = \frac{2}{11}$ and $\beta = \frac{42}{11}$.

Problem 2. Here, we consider a singularly perturbed diffusion coefficient:

$$A_2(x,y) = \operatorname{diag}(\delta(x,y), \varepsilon^2), \quad \delta(x,y) := \frac{3}{2} + \sin\left(\frac{2\pi x^2}{\varepsilon^{3/2}}\right),$$

where now α becomes small as $\varepsilon \to 0$.

Problem 3. In this case, we choose for each $\tau_i, 1 \le i \le N$, a random coefficient

$$A_3|_{\tau_i} \equiv \gamma I \quad \text{with } \gamma \text{ drawn randomly (uniform) from the set } \{0.01, 0.1, 1, 10, 100\}.$$

The relative approximation errors are averaged over 10 random samples of A_3.

To simply the computations we consider the space S_H of linear finite elements on a Cartesian mesh $\mathcal{G}_H$ consisting of square elements of side length H. From finite element theory (cf., e.g., [14, p. 539], [17, Corollary 5.3]) we know that the choice $H = O(\varepsilon)$ in

$$\text{Find } u_H \in S_H \quad \text{such that} \quad a(u_H, v) = F(v) \qquad \forall v \in S_H$$

leads to a solution u_H which has a significantly reduced accuracy compared to problems with smooth coefficients—in other words, *just* resolving the geometry and using a standard linear finite element space is not enough to obtain optimal order convergence of the Galerkin solution.

Our goal is to construct for the "coarse" mesh $\mathcal{G}_H$, $H = O(\varepsilon)$, an "optimal" finite element space $\tilde{S}_H$ by using the Galerkin discretization on an overrefined mesh $\mathcal{G}_h$ with mesh width $h = O\left(H^2\right)$. We use a subscript H if a quantity is related to the mesh $\mathcal{G}_H$ and write $\omega_{i,H}$, $b_{i,H}$, n_H instead of ω_i, b_i, n.

The term "optimal" is understood here in the sense that we construct as follows a low-dimensional subspace $\tilde{S}_H \subset S_h$ as the span of some basis functions $\tilde{b}_i$, i.e.,

$$\tilde{S}_H = \operatorname{span}\left\{\tilde{b}_i \in S_h \mid 1 \le i \le N_H \text{ and } \operatorname{supp}\tilde{b}_i \subset \omega_{i,H}\right\},$$

for some $N_H = O(n_H)$ which *locally* has an optimal approximation behavior with respect to the relative L^2 error: Let $u_{h,j}$ denote the Galerkin solution of (3) with right-hand side f_j (cf. (20)) for the piecewise linear finite element space S_h on the overrefined mesh $\mathcal{G}_h$. We choose the mesh cell

$$\tau := \left(\frac{1}{2} - H, \frac{1}{2}\right) \times \left(\frac{1}{2} - H, \frac{1}{2}\right)$$

exemplarily to construct the local space $\tilde{S}_H(\tau) = \tilde{S}_H\big|_\tau$ as the minimizer in

$$\inf_{\substack{V \subset S_h|_\tau \\ \dim V = 4}} \sup_{1 \le j \le 20} \inf_{v \in V} \frac{\|u_{h,j} - v\|_{L^2(\tau)}}{\|u_{h,j}\|_{L^2(\tau)}}.$$

The local space $\tilde{S}_H(\tau)$ can be obtained by a singular value decomposition of the fine grid solutions as explained below.

Let the vector c_j be the coefficient vector of the Galerkin solution $u_{h,j} \in S_h$ with respect to the standard "hat" basis on the fine mesh $\mathcal{G}_h$ for the right-hand side f_j. We denote by $c_{j,\tau}$ the restriction of c_j to the values at vertices of the cell τ. We define the matrix A columnwise,

$$A := [c_{1,\tau} \mid \cdots \mid c_{20,\tau}],$$

and compute the left and right singular vectors v_j, w_j as well as the corresponding singular values $\sigma_{\tau,j}$ of A. By the inequality

$$\left\| A - \sum_{r=1}^{k} v_r \sigma_{\tau,r} w_r^{\mathsf{T}} \right\|_2 \leq \sigma_{\tau,k+1}$$

we conclude that

$$\|c_{j,\tau} - \tilde{c}_{j,\tau}\|_2 \leq \sigma_{\tau,k+1}, \quad \text{where} \quad \tilde{c}_{j,\tau} := \sum_{r=1}^{k} v_r \sigma_{\tau,r} (w_r)_j$$

holds. This says that the first k left singular vectors v_r of A define a k-dimensional space $S_\tau := \operatorname{span}\{v_r \mid r = 1, \ldots, k\}$ for the cell τ such that we can approximate the τ-parts of the coefficients of all solutions u_j in S_τ up to an error of size $\sigma_{\tau,k+1}$. Note that quadratic convergence with respect to the relative $L^2(\tau)$-norm is equivalent to quadratic convergence of the ratio $\sigma_{\tau,5}/\sigma_{\tau,1}$ as a function of H.

5.2. Decay of singular values. In this section, we compute the singular values for the cell τ and investigate their decay behavior for our model problems. In the first experiment (Problem 1) we consider the diffusion coefficient $A_{1,\varepsilon}$ as in (21) with the choice $\varepsilon = H$.

For comparison we used standard piecewise linear finite elements and a sufficient quadrature (regular refinement to fine-scale) for the setup of the stiffness matrix in order to compute the approximate solution u_{lin}.

The results in Table 1 show that for Problem 1 with fine-scale oscillations the optimal shape functions preserve the quadratic convergence rate (cf. Figures 1 and 2 for plots of the four shape functions), whereas the piecewise linear finite elements are not sufficient for quadratic or even linear convergence rates.

In the case of the singularly perturbed Problem 2 we observe at least linear convergence in Table 2. Note that the coefficient A in this case is not uniformly elliptic as $H \to 0$, and, hence, the assumptions for the theory are violated. This was the only example in which the convergence rates for the "optimal" shape functions are found to be clearly less than quadratic.

In the last experiment (Problem 3), the results in Table 3 show that, even for this medium-contrast random coefficient, the rate of convergence is quadratic.

The numerical experiments were conducted with the HLIB software library (http://www.hlib.org). We conclude that for all test examples only four (properly selected) basis functions per cell preserve the convergence with respect to the relative L^2-norm even for cases where the diffusion coefficient is rather complicated.

TABLE 1

Problem 1. *Convergence rates $\sigma_{\tau,5}/\sigma_{\tau,1}$ for the optimal shape functions and standard piecewise linear shape functions for a nonperiodic oscillating coefficient.*

$H =$	Optimal basis	Ratio	Pw. lin.	Ratio
0.25	1.59_{e-2}		4.65_{e-2}	
0.125	7.76_{e-3}	2.04	2.75_{e-2}	2.91
0.0625	1.82_{e-3}	4.26	1.83_{e-2}	2.18
0.03125	4.44_{e-4}	4.10	1.59_{e-2}	1.15

TABLE 2

Problem 2. *Convergence rates $\sigma_{\tau,5}/\sigma_{\tau,1}$ for the optimal shape functions and standard piecewise linear shape functions for a singularly perturbed problem.*

$H =$	Optimal basis	Ratio	Pw. lin.	Ratio
0.25	5.75_{e-3}		1.93_{e-1}	
0.125	1.97_{e-3}	2.92	2.06_{e-1}	0.94
0.0625	9.04_{e-4}	2.18	2.23_{e-1}	0.92
0.03125	2.93_{e-4}	3.09	2.36_{e-1}	0.94

TABLE 3

Problem 3. *Convergence rates for the optimal shape functions for a random diffusion coefficient (average, minimal, and maximal error from* 10 *samples).*

$H =$	Avg	Ratio	Min	Max
0.25	3.10_{e-1}		1.80_{e-1}	4.06_{e-1}
0.125	7.80_{e-2}	3.97	6.20_{e-2}	9.17_{e-2}
0.0625	1.84_{e-2}	4.24	1.63_{e-2}	2.04_{e-2}
0.03125	4.33_{e-3}	4.25	4.22_{e-3}	4.48_{e-3}

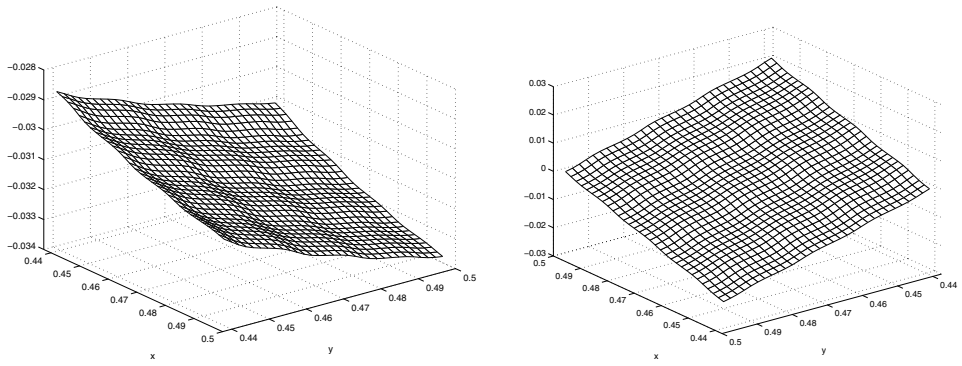

FIG. 1. *The first and second shape functions for the oscillatory coefficient of Problem* 1.

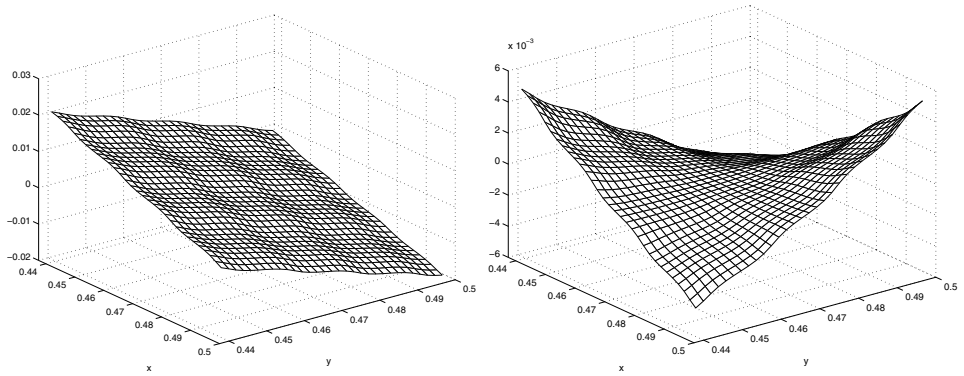

FIG. 2. *The third and fourth shape functions for the oscillatory coefficient of Problem* 1.

Acknowledgment. Part of this work was carried out during a stay of the second and third authors at the Mathematische Forschungsinstitut Oberwolfach, Germany. This support is gratefully acknowledged.

REFERENCES

[1] W. AQUINO, J. C. BRIGHAM, C. J. EARLS, AND N. SUKUMAR, *Generalized finite element method using proper orthogonal decomposition*, Internat. J. Numer. Methods Engrg., 79 (2009), pp. 887–906.

[2] I. BABUŠKA, G. CALOZ, AND J. E. OSBORN, *Special finite element methods for a class of second order elliptic problems with rough coefficients*, SIAM J. Numer. Anal., 31 (1994), pp. 945–981.

[3] I. BABUŠKA AND J. E. OSBORN, *Generalized finite element methods: Their performance and their relation to mixed methods*, SIAM J. Numer. Anal., 20 (1983), pp. 510–536.

[4] I. BABUŠKA AND R. LIPTON, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Mode Simul., 9 (2011), pp. 373–406.

[5] M. BEBENDORF AND W. HACKBUSCH, *Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with L^∞-coefficients*, Numer. Math., 95 (2003), pp. 1–28.

[6] S. BÖRM, *Approximation of solution operators of elliptic partial differential equations by $\mathcal{H}$- and $\mathcal{H}^2$-matrices*, Numer. Math., 115 (2010), pp. 165–193.

[7] P. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1987.

[8] W. DAHMEN, B. FAERMANN, I. GRAHAM, W. HACKBUSCH, AND S. SAUTER, *Inverse inequalities on non-quasiuniform meshes and applications to the mortar element method*, Math. Comp., 73 (2003), pp. 1107–1138.

[9] H. DONG AND S. KIM, *Green's matrices of second order elliptic systems with measurable coefficients in two dimensional domains*, Trans. AMS, 361 (2009), pp. 3303–3323.

[10] R. DU AND P. MING, *Convergence of the heterogeneous multiscale finite element method for elliptic problems with nonsmooth microstructures*, Multiscale Model. Simul., 8 (2010), pp. 1770–1783.

[11] M. GRÜTER AND K.-O. WIDMAN, *The Green function for uniformly elliptic equations*, Manuscripta Math., 37 (1982), pp. 303–342.

[12] A. HENROT, *Extremum Problems for Eigenvalues of Elliptic Operators*, Birkhäuser-Verlag, Basel, 2006.

[13] M. LARSON AND A. MÅLQVIST, *Adaptive variational multiscale methods based on a posteriori error estimation: Energy norm estimates for elliptic problems*, Comput. Methods Appl. Mech. Eng., 196 (2007), pp. 2313–2324.

[14] A. MATACHE AND C. SCHWAB, *Two-scale FEM for homogenization problems*, M2AN. Math. Model. Numer. Anal., 36 (2002), pp. 537–572.

[15] H. Owhadi and L. Zhang, *Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast*, Multiscale Model. Simul., 9 (2011), pp. 1373–1398.
[16] A. Målqvist and D. Peterseim, *Localization of Elliptic Multiscale Problems*, Technical report, arXiv:1110.0692v2, 2011.
[17] D. Peterseim and S. Sauter, *Error Estimates for Finite Element Discretizations of Elliptic Problems with Oscillatory Coefficients*, preprint 13, Institut für Mathematik, Universität Zürich, Zürich, 2010.